Supplemental Material for Morphological Network: How Far Can We Go with Morphological Neurons?

Ranjan Mondal[1]
https://ranjanz.github.io
Sanchayan Santra[2]
https://san-santra.github.io
Soumendu Sundar Mukherjee[3]
https://soumendu041.gitlab.io
Bhabatosh Chanda[3]
bchanda57@gmail.com

[1] Samsung Research
Bangalore
India
[2] Institute for Datability Science
Osaka University
Osaka, Japan
[3] Indian Statistical Institute
Kolkata
India

1 Gradient of Soft maximum

The derivative of soft dilation and erosion operation may be defined as follows.

$$\frac{\delta(\boldsymbol{x}\hat{\oplus}\boldsymbol{s})}{\delta s_k} = \frac{e^{(x_k+s_k)\beta}}{\sum_i e^{(x_i+s_i)\beta}} \tag{1}$$

$$\frac{\delta(\boldsymbol{x}\hat{\ominus}\boldsymbol{s})}{\delta s_k} = \frac{e^{(s_k-x_k)\beta}}{\sum_i e^{(s_i-x_i)\beta}} \tag{2}$$

2 Equivalence of Configurations

In this section, we prove that some of the arrangements of morphological neurons are equivalent and can be approximated by using a fewer number of neurons. To be able to do that, we first prove a simple lemma.

Lemma 1. *Suppose f and g are real-valued functions on $\mathbb{R}^d$. Then $f = g$ if and only if, for all $r \in \mathbb{R}$, one has equality of the sub-level sets:*

$$f^{-1}(-\infty, r] = g^{-1}(-\infty, r].$$

Proof. The "only if" part is trivial. As for the "if" part, note that we have

$$f^{-1}\{r\} = \bigcap_{n\geq 1} f^{-1}(r-1/n, r] = \bigcap_{n\geq 1}(f^{-1}(-\infty, r] \setminus f^{-1}(-\infty, r-1/n]).$$

© 2022. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

The same goes for g, and so, by our hypothesis,

$$f^{-1}\{r\} = g^{-1}\{r\} \quad \text{for all } r \in \mathbb{R}.$$

Therefore, for any $x \in \mathbb{R}^d$, we have $x \in g^{-1}\{g(x)\} = f^{-1}\{g(x)\}$, or, in other words, $f(x) = g(x)$. □

Theorem 1. *If we denote $D_{m_1}E_{m_2}$ as a layer with m_1 dilation neurons and m_2 erosion neurons and L as a linear combination layer, the following may be said about their different configurations.*

(i) The architecture $D_{m_1}E_0 \to D_{m_2}E_0 \to \cdots \to D_{m_\ell}E_0$ consisting only of dilation layers is equivalent to the architecture $D_{m_\ell}E_0$ with a single dilation layer. A similar statement is true if one considers architectures with only purely erosion layers.

(ii) The architecture $D_1E_1 \to D_1$ is not equivalent to D_1E_0. Similarly, it is not equivalent to D_0E_1, and, consequently, the architectures $D_1E_1 \to D_1E_1$ and D_1E_1 are inequivalent.

(iii) The architecture $D_1E_1 \to D_1 \to L$ is not equivalent to $D_1E_0 \to L$.

(iv) The architecture $D_2E_0 \to D_0E_2 \to D_1$ is not equivalent to $D_2E_0 \to D_1$.

Proof. (i) Let $x \in R^d$ be the input to the network. Let there be two networks N_1 and N_2. Let there be m_1 and m_2 dilated neurons in, respectively, the first and the second layers of Network N_1. Let the parameters of the network N_1 in the first layer and 2nd layer are $w^1 \in R^{d\times m_1}$ and $w^2 \in R^{l_1\times l_2}$ respectively. Whereas let there is only a single layer with m_1 number of dilated neurons in network N_2 and the parameters are denoted as $u \in R^{d\times m_2}$. Let $f(x) \in R^{m_2}$ and $g(x) \in R^{m_2}$ are the output from the last layer of network N_1 and N_2 respectively.

For Network N_1

$$y_j = \max_i(x_i + w^1_{i,j}) \quad \forall j \in \{1,2,..m_1\} \tag{3}$$

$$f_k(x) = \max_j(y_j + w^2_{j,k}) \quad \forall j,k \tag{4}$$

For network N_2

$$g_k(x) = \max_j(x_j + u^2_{j,k}) \qquad \forall k,j \tag{5}$$

Let

$$S^k_f = \{x \mid f_k(x) \le e_k; e_k \in R\} \tag{6}$$

$$S^k_g = \{x \mid f_k(x) \le e_k; e_k \in R\} \tag{7}$$

For Network N_1

$$f_k(x) \le e_k; \qquad \forall k \tag{8}$$

$$y_i + w^2_{i,j} \le e_k \qquad \forall k,j \tag{9}$$

$$y_i \le e_k - w^2_{i,j} \qquad \forall k,j \tag{10}$$

From equation 3 and equation 10 we get

$$\max_i (x_i + w^1_{i,j}) \leq e_k - w^2_{i,j} \qquad \forall k, j \tag{11}$$

$$x_i + w^1_{i,j} \leq e_k - w^2_{i,j} \qquad \forall k, j, i \tag{12}$$

$$x_i \leq e_k - w^2_{i,j} - w^1_{i,j} \qquad \forall k, j, i \tag{13}$$

Which means

$$x_i \leq \min_j (e_k - w^2_{i,j} - w^1_{i,j}) \qquad \forall k, i \tag{14}$$

$$x_i \leq e_k - \max_j (w^2_{i,j} + w^1_{i,j}) \qquad \forall k, i \tag{15}$$

For network N_2

$$g_k(x) = \max_j (x_j + u^2_{j,k}) \tag{16}$$

$$x_i \leq (e_k - u_{i,k}) \forall k, i \tag{17}$$

To hold the set S^k_g is equal to S^k_f to $\forall k$

$$u_{i,k} = \max_j (w^2_{i,j} + w^1_{i,j}) \forall i, k \tag{18}$$

Hence, from Lemma 1, given a parameter w^1 and w^2 of and 2 layer network N_1, there exist a equivalent single-layer network N_2 with dilated neurons u which can represent the same function.From the equation 18 we can see the parameters of the single-layer network can be constructed considering the longest path from input to output. Recursively we can say it holds for multiple layers. A similar argument can be given in the case of erosion layers. Hence $D_{m_1}E_0 \rightarrow D_{m_2}E_0 \rightarrow \cdots \rightarrow D_{m_\ell}E_0$ proved

(ii) For simplicity, we will assume 2-dimensional input. Suppose that the outputs from the first layer are $f_1(x,y)$ and $g_1(x,y)$ where f_1 is the output of a dilation neurone and g_1 is the output of an erosion neurone (see Figure 1). We write

$$f_1(x,y) = \max\{x+a, y+b\}, \quad g_1(x,y) = \min\{x+c, y+d\}.$$

After the second layer consisting of a single dilation neurone, we get the output

$$f_2(x,y) = \max\{f_1 + a_1, g_1 + b_1\}.$$

Note that

$$\begin{aligned}
f_2(x,y) \leq e &\iff f_1 + a_1 \leq e \text{ and } g_1 + b_1 \leq e \\
&\iff f_1 \leq e - a_1 \text{ and } g_1 \leq e - b_1 \\
&\iff (x + a \leq e - a_1 \text{ and } y + b \leq e - a_1) \text{ and } (x + c \leq e - b_1 \text{ or } y + d \leq e - b_1) \\
&\iff (x,y) \in (-\infty, \gamma_1] \times (-\infty, \gamma_2] \cap ((-\infty, \gamma_3] \times \mathbb{R} \cup \mathbb{R} \times (-\infty, \gamma_4]) \\
&\iff (x,y) \in (-\infty, \gamma_1 \wedge \gamma_3] \times (-\infty, \gamma_2] \cup (-\infty, \gamma_1] \times (-\infty, \gamma_2 \wedge \gamma_4].
\end{aligned}$$

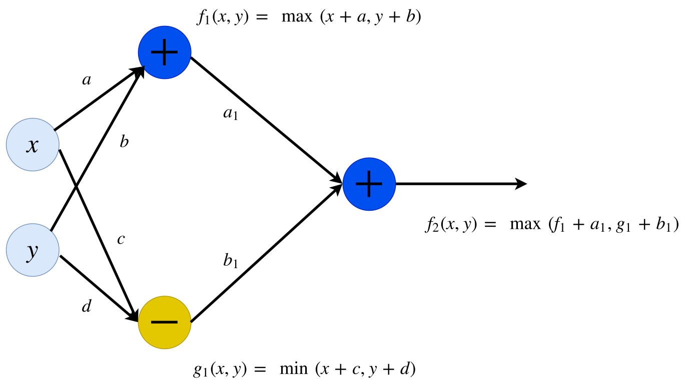


Figure 1: A network of architecture $D_1E_1 \to D_1$

Note that $\gamma_1 \leq \gamma_3 \iff a_1 + a \geq b_1 + c$ and $\gamma_2 \leq \gamma_4 \iff a_1 + b \geq b_1 + d$. Therefore, if $a_1 + a \geq b_1 + c$ and $a_1 + b \geq b_1 + d$, then

$$f_2^{-1}(-\infty, e] = (-\infty, \gamma_1] \times (-\infty, \gamma_2].$$

Thus in this case f_2 can be realized in the architecture D_1E_0.

If, however, $a_1 + a < b_1 + c$ and $a_1 + b < b_1 + d$, then

$$f_2^{-1}(-\infty, e] = (-\infty, \gamma_3] \times (-\infty, \gamma_2] \cup (-\infty, \gamma_1] \times (-\infty, \gamma_4],$$

which is not realizable as the sublevel set of a function of D_1E_0 architecture.

(iii) The proof is a simple modification of the proof of (ii). For $\alpha > 0$,

$$\begin{aligned}
\alpha f_2(x,y) \leq e &\iff f_1 + a_1 \leq \frac{e}{\alpha} \text{ and } g_1 + b_1 \leq \frac{e}{\alpha} \\
&\iff f_1 \leq \frac{e}{\alpha} - a_1 \text{ and } g_1 \leq \frac{e}{\alpha} - b_1 \\
&\iff (x + a \leq \frac{e}{\alpha} - a_1 \text{ and } y + b \leq \frac{e}{\alpha} - a_1) \\
&\quad \text{and } (x + c \leq \frac{e}{\alpha} - b_1 \text{ or } y + d \leq \frac{e}{\alpha} - b_1) \\
&\iff (x,y) \in (-\infty, \gamma_1] \times (-\infty, \gamma_2] \cap ((-\infty, \gamma_3] \times \mathbb{R} \cup \mathbb{R} \times (-\infty, \gamma_4]) \\
&\iff (x,y) \in (-\infty, \gamma_1 \wedge \gamma_3] \times (-\infty, \gamma_2] \cup (-\infty, \gamma_1] \times (-\infty, \gamma_2 \wedge \gamma_4].
\end{aligned}$$

Note that $\gamma_1 \leq \gamma_3 \iff a_1 + a \geq b_1 + c$ and $\gamma_2 \leq \gamma_4 \iff a_1 + b \geq b_1 + d$.

Therefore, if $a_1 + a \geq b_1 + c$ and $a_1 + b \geq b_1 + d$, then

$$(\alpha f_2)^{-1}(-\infty, e] = (-\infty, \gamma_1] \times (-\infty, \gamma_2].$$

Thus in this case f_2 can be realized in the architecture $D_1E_0 \to L$.

If, however, $a_1 + a < b_1 + c$ and $a_1 + b < b_1 + d$, then

$$(\alpha f_2)^{-1}(-\infty, e] = (-\infty, \gamma_3] \times (-\infty, \gamma_2] \cup (-\infty, \gamma_1] \times (-\infty, \gamma_4],$$

which is not realizable as the sublevel set of a function of $D_1E_0 \to L$ architecture.

Sub-level sets of the $D_1E_0 \to L$ architecture. For $\beta > 0$

$$\beta \max\{x+u, y+v\} \leq e \iff x \leq \frac{e}{\beta} - u \text{ and } y \leq \frac{e}{\beta} - v. \tag{19}$$

Equating $\frac{e}{\beta} - u = \frac{e}{\alpha} - a_1 - a$, $\frac{e}{\beta} - v = \frac{e}{\alpha} - a_1 - b$, we can see that one can take $\beta = \alpha, u = a + a_1, v = b + a_1$ to realize the function αf_2 in the $D_1E_0 \to L$ architecture.

(iv) It can be proved in the same way as (ii) □

3 Proof of Proposition 1: Morphological block as a sum of hinge functions

Proposition 1. *The function computed by a single morphological block with n dilation and m erosion neurons followed by a linear combination layer computes $\mathcal{M}(\boldsymbol{x})$, which is a sum of multi-order hinge functions.*

Proof. As defined in the main paper the computed $\mathcal{M}(\boldsymbol{x})$ has the following form.

$$\mathcal{M}(\boldsymbol{x}) = \sum_{i=1}^{n} \omega_i^+ z_i^+ + \sum_{j=1}^{m} \omega_j^- z_j^-, \tag{20}$$

where z_i^+ and z_j^- are the output of i^{th} dilation neuron and j^{th} erosion neuron, respectively and ω_i^+ and ω_j^- are the weights of the the linear combination layer. Replacing the z_i^+ and z_j^- with their expression, the equation becomes the following.

$$\mathcal{M}(\boldsymbol{x}) = \sum_{i=1}^{n} \omega_i^+ \max_k\{x'_k + s_{ik}^+\} + \sum_{i=1}^{m} -\omega_i^- \max_k\{s_{ik}^- - x'_k\}, \tag{21}$$

where s_{ik}^+ and s_{ik}^- denote the k^{th} component of the i^{th} structuring element of dilation and erosion neurons, respectively. The above equation can be further expressed in the following form,

$$\mathcal{M}(\boldsymbol{x}) = \sum_{i=1}^{n} \alpha_i^+ \max_k\{\theta i^+ x'_k + \rho_{ik}^+\} + \sum_{i=1}^{m} \alpha_i^- \max_k\{\theta_i^- x'_k + \rho_{ik}^-\}, \tag{22}$$

Where θ_i^+, θ_i^-, ρ_{ik}^+ and ρ_{ik}^- are defined in the following way

$$\theta_i^+ = \begin{cases} \omega_i^+ & \text{if } \omega_i^+ \geq 0 \\ -\omega_i^+ & \text{if } \omega_i^+ < 0 \end{cases} \qquad \theta_i^- = \begin{cases} -\omega_i^- & \text{if } \omega_i^- \geq 0 \\ \omega_i^- & \text{if } \omega_i^- < 0 \end{cases}$$

$$\rho_{ik}^+ = \begin{cases} s_{ik}^+ \omega_i^+ & \text{if } \omega_i^+ \geq 0 \\ -s_{ik}^+ \omega_i^+ & \text{if } \omega_i^+ < 0 \end{cases} \qquad \rho_{ik}^- = \begin{cases} s_{ik}^- \omega_i^- & \text{if } \omega_i^- \geq 0 \\ -s_{ik}^- \omega_i^- & \text{if } \omega_i^- < 0 \end{cases}$$

$$\alpha_i^+ = \begin{cases} 1 & \text{if } \omega_i^+ \geq 0 \\ -1 & \text{if } \omega_i^+ < 0 \end{cases} \qquad \alpha_i^- = \begin{cases} -1 & \text{if } \omega_i^- \geq 0 \\ 1 & \text{if } \omega_i^- < 0 \end{cases}$$

Now, without any loss of generality, we can write equation 22 as follows

$$\mathcal{M}(\boldsymbol{x}) = \sum_{i=1}^{m+n} \alpha_i \max_k(\theta_i x'_k + \rho_{ik}) \tag{23}$$

where

$$\theta_i = \begin{cases} \theta i^+ & \text{if } i \leq n \\ \theta_{i-n}^- & \text{if } n < i \leq m+n \end{cases}$$

$$\rho_{ik} = \begin{cases} \rho_{ik}^+ & \text{if } i \leq n \\ \rho_{(i-n)k}^- & \text{if } n < i \leq m+n \end{cases}$$

$$\alpha_i = \begin{cases} \alpha_i^+ & \text{if } i \leq n \\ \alpha_{(i-n)}^- & \text{if } n < i \leq m+n \end{cases}$$

Finally, we can rewrite equation 23 as

$$\mathcal{M}(\boldsymbol{x}) = \sum_{i=1}^{l} \alpha_i \phi_i(\boldsymbol{x}), \tag{24}$$

where $l = m+n$, $\alpha_i \in \{1, -1\}$ and $\phi_i(\boldsymbol{x})$'s are of the following form

$$\phi_i(\boldsymbol{x}) = \max_k (\boldsymbol{v}_{ik}^T \boldsymbol{x}' + \rho_{ik}), \tag{25}$$

with

$$v_{ikt} = \begin{cases} \beta_i & \text{if } t = k \\ 0 & \text{if } t \neq k \end{cases} \tag{26}$$

In equation 25, $\boldsymbol{v}_{ik}^T \boldsymbol{x}' + \rho_{ik}$ is affine and $\alpha_i \phi_i(\boldsymbol{x})$ is a d-order hinge function. Hence $\sum_{i=1}^{l} \alpha_i \phi_i(\boldsymbol{x})$ i.e., $\mathcal{M}(\boldsymbol{x})$ represents sum of multi-oder hinge function. □

4 Number of Hyperplanes

Since a morphological block computes a sum of hinge functions, it can potentially learn a large number of hyperplanes. The function $\mathcal{M}(\boldsymbol{x})$ learned by a single-layer Morphological network may also be expressed in the following form:

$$\mathcal{M}(\boldsymbol{x}) = \sum_{i=1}^{l} \alpha_i \max_k \{\theta_k x_k + \rho_{ik}\}, \tag{27}$$

where α_i, θ_k, $\rho_{ik} \in \mathbb{R}$. We see that $\mathcal{M}(\boldsymbol{x})$ is a sum of l functions, each of which computes max over the linearly transformed elements of $\boldsymbol{x}$. Since the max is computed over the (transformed) elements of $\boldsymbol{x}$, each max operation selects only one element of $\boldsymbol{x}$. So, the computed $\mathcal{M}(\boldsymbol{x})$ may not contain all the elements of $\boldsymbol{x}$ and the index (k) of the selected element varies depending on the input and the structuring element. However, if $l > d$, $\mathcal{M}(\boldsymbol{x})$ may contain all the elements of $\boldsymbol{x}$. So equation 27 can be rewritten as

$$\mathcal{M}(\boldsymbol{x}) = \alpha_1(\theta_1 x_{k_1} + \rho_{1k_1}) + \alpha_2(\theta_2 x_{k_2} + \rho_{2k_2}) + \cdots + \alpha_l(\theta_l x_{k_l} + \rho_{lk_l}). \tag{28}$$

where x_{k_i} represents any one of the $d+1$ elements of $\boldsymbol{x}$ selected by i-th neuron by max operation depending on structuring element $\boldsymbol{s}_i$. So each x_{k_i} is chosen from $d+1$ elements.

Therefore, depending on which element of $\boldsymbol{x}$ gets selected by each neuron, $\mathcal{M}(\boldsymbol{x})$ forms one of the $(d+1)^l - 1$ hyperplanes. The -1 occurs in the number of hyperplanes because on one occasion only limiter or bias is selected. Note that some of these hyperplanes must be parallel to some axes. For $\mathcal{M}(\boldsymbol{x})$ to form a hyperplane that is not parallel to any of the axes, all elements of $\boldsymbol{x}$ must get selected by some max functions or other. This occurs in $d! \times \binom{l}{d}$ ways. The remaining $l-d$ number of elements x_{k_i}'s are repeat selection by some functions. So, there can be almost $d! \times \binom{l}{d} \times (d+1)^{l-d}$ hinging hyperplanes that are not parallel to any of the axes.

5 One morphological block and function approximation

A single morphological block represents a sum of hinge functions. However, it is not clear if all hinge functions can be represented by a single morphological block. In a numerical study, we have tried to approximate the hinge function $\max(x+y,0)$ using a single morphological block by varying the number of dilation/erosion neurons. We have generated values of the function in the square $[-5,5] \times [-5,5]$, and trained the network with mean squared error (MSE) loss. In Figure 2, we have plotted the MSE loss (after convergence) against the number of morphological neurons used. It is seen that a single morphological block is unable to reduce the error unless we use additional bias or limiter terms in the morphological neurons. However, we do not know theoretically if having additional bias terms in morphological operations help in universal approximation.

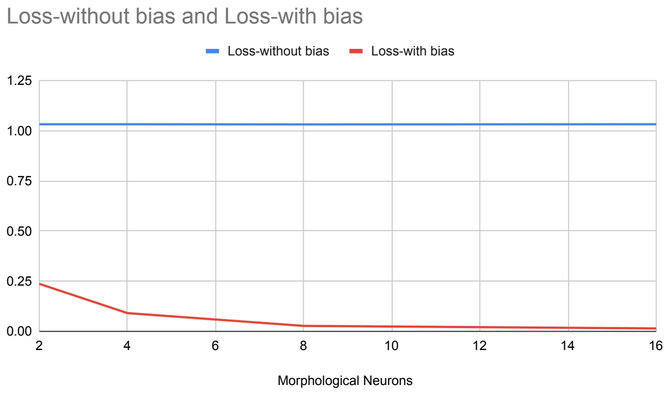


Figure 2: Graph of approximation loss with varying morphological neurons in a single morphological block.

6 Universal Approximation by two Morphological blocks

Here we have shown that two sequential Morphological blocks can approximate any continuous functions. First, we have shown that any hyperplane can be represented by a single morphological block. After that, we have shown the universal approximation using two morphological blocks.

Lemma 2. *Let K be a compact subset of $\mathbb{R}^d$. Then, over K, any hyperplane $w^\top \boldsymbol{x} + b$ can be represented as an affine combination of d dilation neurons which only depend on K.*

Proof. Since we are in a compact set, there exists $C > 0$ such that $|x_\ell| \le C$ for any $1 \le \ell \le d$. Where x_ℓ is each element of $\boldsymbol{x}$. Take

$$s_\ell = -3C\mathbf{1}_d + 3Ce_{\ell,d}, 1 \le \ell \le d,$$

where $\mathbf{1}_d$ is the vector of all ones and $e_{\ell,d}$ is the ℓ-th unit vector in $\mathbb{R}^d$. Then all but the ℓ-th coordinate of s_ℓ are $-3C$, while the ℓ-th coordinate is 0. Then note that, for any $\boldsymbol{x} \in K$, and $1 \le \ell \le d$,

$$\begin{aligned} x_\ell + s_{\ell,\ell} = x_\ell \ge -C > -2C = C - 3C \\ \ge x_j - 3C = x_j + s_{\ell,j}, \end{aligned}$$

for any $j \ne \ell$. It follows that for any $x \in K$, and $1 \le \ell \le d$,

$$\boldsymbol{x} \oplus s_\ell = x_\ell.$$

Now given any hyperplane $w^\top \boldsymbol{x} + b$, we can express it exactly as a linear combination of dilation neurons over K:

$$w^\top \boldsymbol{x} + b = \sum_{\ell=1}^{d} w_\ell x_\ell + b = \sum_{\ell=1}^{d} w_\ell (\boldsymbol{x} \oplus s_\ell) + b.$$

This completes the proof. □

Lemma 3 (lemma 1 of main paper)**.** *Any linear combination of hinge functions $\sum_{i=1}^{m} \alpha_i h^{(k_i)}(\boldsymbol{x})$ can be represented over any compact set K as a two sequential morphological block consisting of dilation neurons only.*

Proof. Let $B = \max_{1 \le i \le m} \sup_{\boldsymbol{x} \in K} |h^{(k_i)}(\boldsymbol{x})|$. We now give the architecture of the desired Morph-Net.

1. The first dilation-erosion layer has exactly d dilation neurons given by $\boldsymbol{x} \oplus s_\ell, 1 \le \ell \le d$.
2. The first linear combination layer has $k = \sum_{i=1}^{m} (k_i + 1)$ neurons, with the i-th block of $(k_i + 1)$ neurons outputting the constituent hyperplanes of $h^{(k_i)}(\boldsymbol{x})$. This can be done by Lemma 2.
3. The second dilation-erosion layer just has m dilation neurons, each outputting a hinge function. The ℓ-th neuron is constructed as follows: Write any $\boldsymbol{y} \in \mathbb{R}^k$ as $(\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_m^\top)^\top$ where $\boldsymbol{y}_j = (y_{j,1}, \ldots, y_{j,k_j+1})^\top$. We want the output of the ℓ-th neuron to be $\max_{1 \le v \le k_\ell + 1} y_{\ell,v}$. So we take $\boldsymbol{t}_\ell = (\boldsymbol{t}_{\ell,1}^\top, \ldots, \boldsymbol{t}_{\ell,m}^\top)^\top$, where $\boldsymbol{t}_{\ell,j} = -3B\mathbf{1}_{k_j+1}$ for $j \ne \ell$, and $\boldsymbol{t}_{\ell,\ell} = \mathbf{0}_{k_\ell+1}$. Then, for any $j \ne \ell$, $1 \le u \le k_j + 1$, and $1 \le v \le k_\ell + 1$, we have

$$\begin{aligned} y_{j,u} + t_{\ell,j,u} = y_{j,u} - 3B &\le B - 3B \\ &= -2B \\ &< -B \\ &\le y_{\ell,v} = y_{\ell,v} + t_{\ell,\ell,v}. \end{aligned}$$

   It follows that $\boldsymbol{y} \oplus \boldsymbol{t}_\ell = \max_{1 \le v \le k_\ell + 1} y_{\ell,v}$. With this construction, the outputs of the second dilation-erosion layer are the m numbers $h^{(k_i)}(\boldsymbol{x})$.

4. The second linear combination layer just has a single neuron that combines the outputs of the previous layer in the desired way:

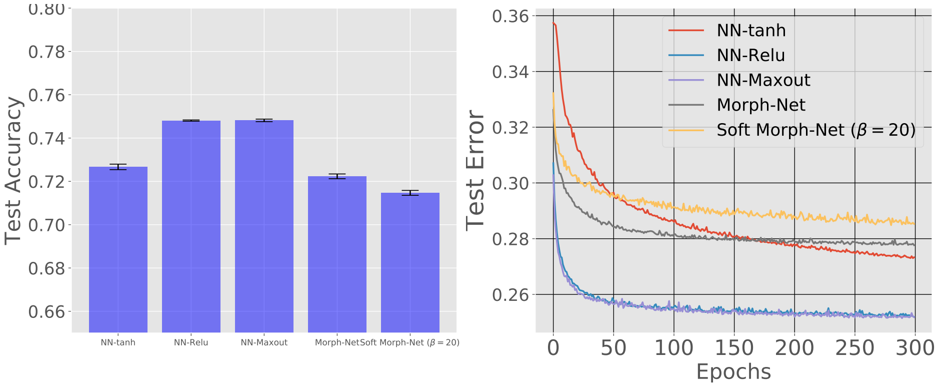
$$z \mapsto \sum_{i=1}^{m} \alpha_i z_i.$$

This completes the proof. □

7 Results

7.1 HIGGS Dataset

Here also show some results on the Higgs dataset. Here the performance of the morphological block is not so good, but this could provide some hint towards improving the performance.

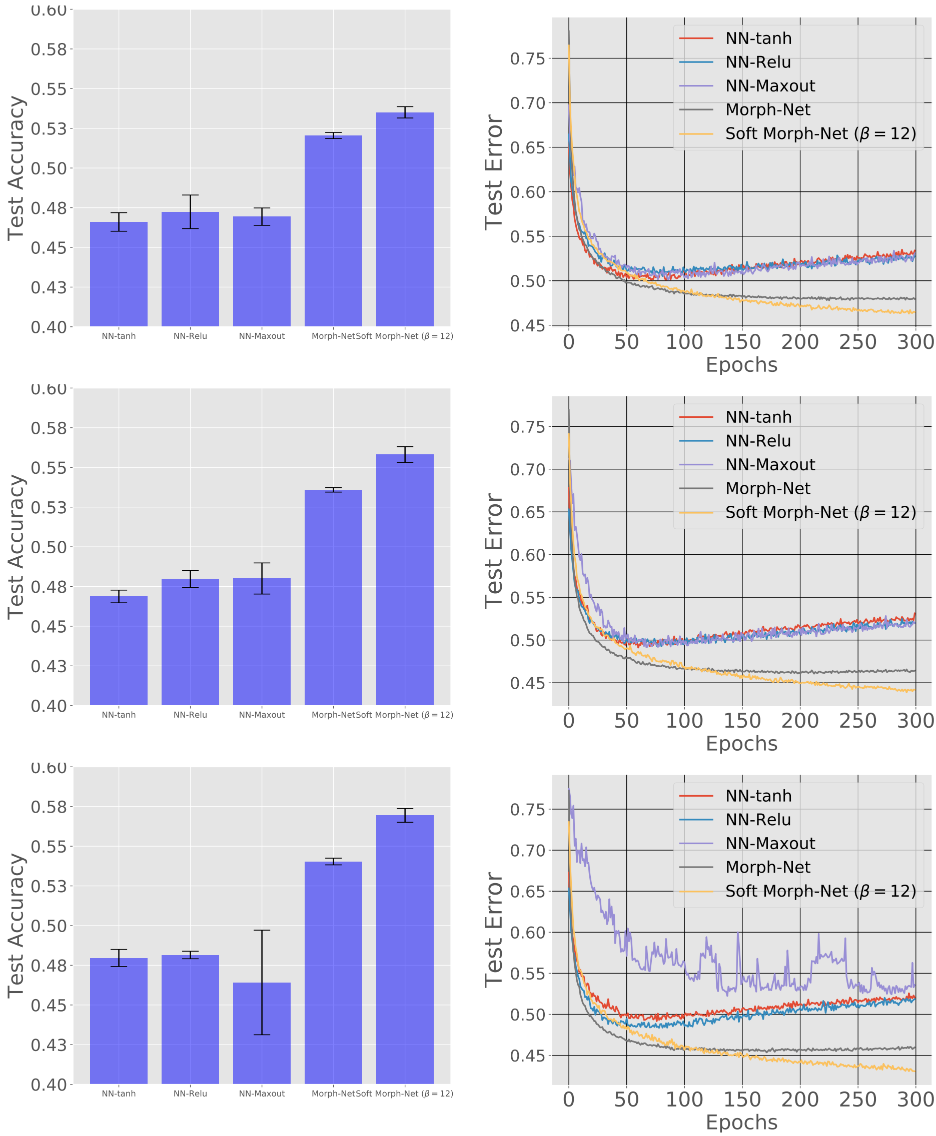


(a) input (b) Morph-Net

Figure 3: Results on Higgs dataset

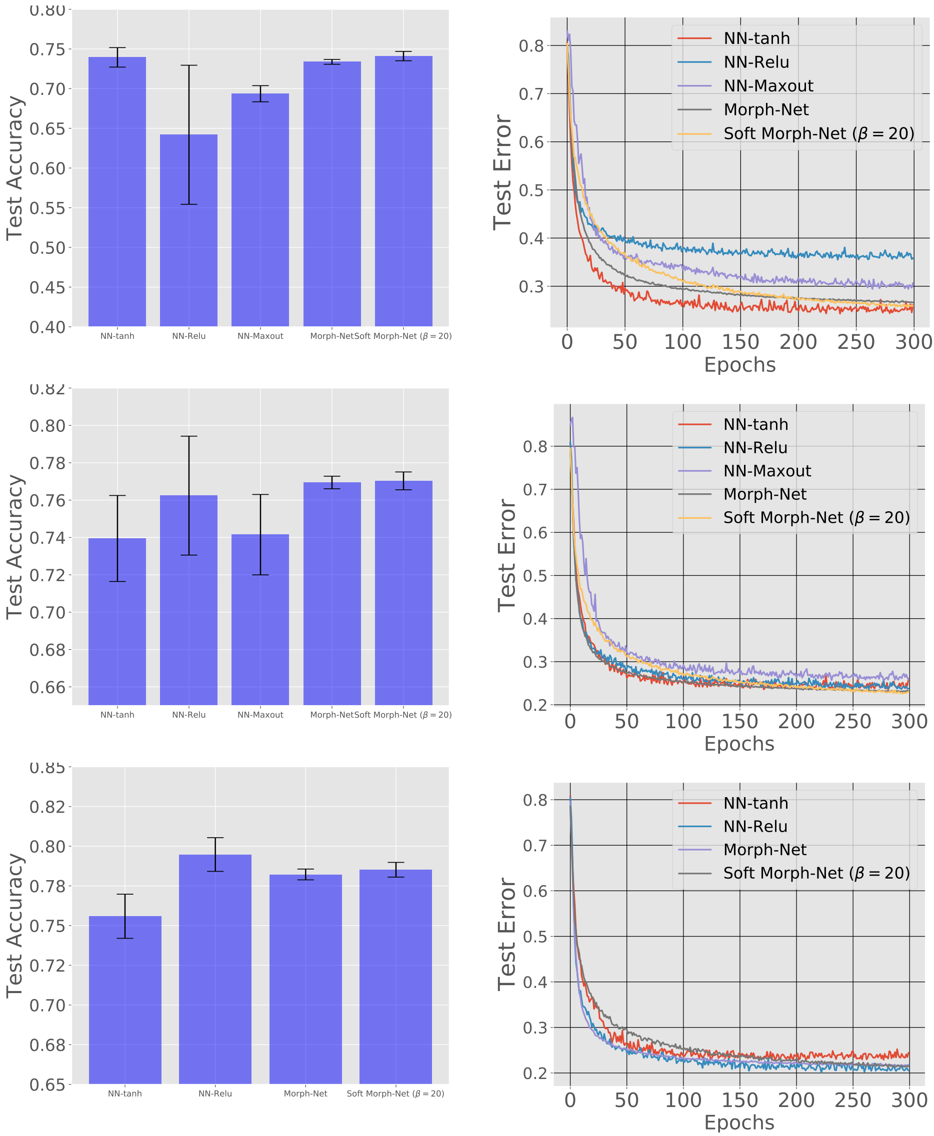
7.2 CIFAR10 and SVHN

A few results of CIFAR10 and SVHN dataset by varying number of neurons is given in figure 4 and figure 5.



(a) Bar graph (b) Test error vs epochs

Figure 4: Results on CIFAR10 dataset, varying number of neurons in the hidden layer: l=200 (1st row), l=400 (2nd row), l=600 (3rd row)



(a) Bar graph (b) Test error vs epochs

Figure 5: Results on SVHN dataset, varying number l=200 (1st row), l=400 (2nd row), l=600 (3rd row)